(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 March 2002 (21.03.2002)

PCT

(10) International Publication Number
WO 02/22873 A1

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/SE01/01691

(22) International Filing Date: 1 August 2001 (01.08.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0003286-2 15 September 2000 (15.09.2000) SE

(71) Applicants and
(72) Inventors: GYLLENSTEN, Ulf [SE/SE]; Malma Backe 7H, S-756 47 Uppsala (SE). INGMAN, Max [SE/SE]; Täljstensvägen 10B, S-752 40 Uppsala (SE). ALLEN, Marie [SE/SE]; Malma Backe 4D, S-756 47 Uppsala (SE). ANDREASSON, Hanna [SE/SE]; Malma Backe 3H, S-756 47 Uppsala (SE).

(74) Agent: DR LUDWIG BRANN PATENTBYRÅ AB; Box 1344, S-751 43 Uppsala (SE).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AT (utility model), AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, CZ (utility model), DE, DE (utility model), DK, DK (utility model), DM, DZ, EC, EE, EE (utility model), ES, FI, FI (utility model), GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK (utility model), SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *with amended claims*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/22873 A1

(54) Title: METHOD AND KIT FOR DETERMINING HUMAN GEOGRAPHIC OR POPULATION ORIGIN

(57) Abstract: The present invention is within the fields of human origins, medicine and evolutionary biology. More precisely, the invention relates to a method and kit for determining the geographic or population origin of a human based on mitochondrial DNA sequences and specifically to the use of mitochondrial DNA variants (polymorphisms) from the complete human mitochondrial genome. This information is employed in the comparison of biological samples with samples of known origin or with a database of mitochondrial genome sequences.

Method and kit for determining human geographic or population origin

Field of invention

The present invention is within the fields of human origins, medicine and evolutionary biology. More precisely, the invention relates to a method and kit for determining the geographic or population origin of a human based on mitochondrial DNA sequences and specifically to the use of mitochondrial DNA variants (polymorphisms) from the complete human mitochondrial genome. This information is employed in the comparison of biological samples with samples of known origin or with a database of mitochondrial genome sequences.

Background of the invention

Each individual, with the exception of identical twins, has a unique genetic constitution. Therefore, identification of the genetic origin can, in principle, be based on a comparison of the genetic material between e.g. a sample of unknown human origin and reference samples of known origin. Such an analysis is of relevance in maternity and paternity investigations, in immigration cases where the familial relationships is disputed, in medical and evolutionary biology research.

The genetic material (DNA) in humans is found in the cell nucleus (containing over 99% of the material) and the mitochondrion (with less than 1%). The DNA in the nucleus is inherited with 50% from each parent, while the DNA in the mitochondrion (called the mtDNA) is derived solely from the mother (a process denoted maternal inheritance). Several characteristics of the mitochondrial genome makes it different from that in the nucleus. The high mtDNA copy number per cell (1,000 - 10,000 copies/cell) allows for the analysis of materials with limited amounts of, or partly degraded, DNA (Bodenhagen and Clayton, 1974). The higher nucleotide substitution rate of mtDNA relative to most nuclear genes also increases the potential for individual identification (Brown et al. 1982). Finally, mtDNA is inherited uniparentally, through the maternal parent (Hutchinson et al. 1974). Since the mtDNA of siblings and close maternal relatives are expected to be identical, with the exception of new mutations, individuals may be assigned to a maternal lineage. Due to the haploid, or clonal nature, of mtDNA, jumping-PCR (Pääbo et al. 1989), which commonly occurs in degraded DNA, does not cause erroneous results. By contrast, in analysis of nuclear loci, jumping PCR may result

in the presence of chimeric sequences that complicate the interpretation of the allelic phase of polymorphisms.

Studies of human origins have been based on analyses of small segments of either nuclear DNA or mtDNA. The studies of mtDNA have been confined to a small segment of the mtDNA genome, denoted the D-loop (Higuchi et al. 1988, Wilson et al. 1993, Ginther et al. 1992, Hagerberg and Sykes 1989, Stoneking et al. 1991, Holland et al. 1993, Handt et al. 1994, Gill et al. 1994, Allen et al. 1998). No studies including the entire mtDNA for the purpose of determining the maternal lineage and deducing the geographic or population origin have been described. Analysis of the D-loop region have been complicated by the extreme variation in substitution rate between different nucleotide sites (necessitating the exclusion of some sites from the analysis) and the relatively high rate of new mutations (potentially resulting in false inclusions/exclusions). Different tissues of an individual may also show differences in the D-loop mtDNA, rendering comparisons between samples from maternal relatives but of different tissue type ambiguous.

Studies have been performed to estimate probability by which the major ethnic group (or geographic region) of an individual can be estimated based on D-loop sequences (Allen et al. 1998). Using D-loop sequences alone is it impossibile to identify the geographic or population origin of an individual.

Studies of the human mitochondrial molecule have also been carried out through RFLP analysis, providing data on some DNA variants outside the D-loop. These analyses were carried out to study the evolutionary history of the human species, or to identify mutations causing mitochondrial disease and were not performed for the purpose of determining the deducing the geographic and population origin of an individual.

Summary of the invention

The present invention provides a method for determining the geographic and population origin of an individual based on analysis of biological samples. According to the present invention the analysis is based on an analysis of the entire mitochondrial DNA (mtDNA) and comparison of the sample under investigation with that of known origin or with a database of compete mitochondrial genome sequences.

In a first aspect the present invention relates to a method for determining the geographic and population origin of a human comprising the following steps:

a) determining the complete nucleic acid sequence of the mitochondrial genome including polymorphic sites in a sample from a human subject who's origin or identity is studied; and
b) relating the information from step a) to mitochondrial nucleic acid sequence information of known origin.

The body sample referred to above can be derived from body fluid or tissue.

The known information in step b) may be derived from human subjects of known identity (reference subjects). Alternatively, the known information in step b) is derived from a database of the nucleic acid sequence information the complete mitochondrial genome of humans of diverse origin. This may be appropriate in cases where samples from maternal relatives are not available, or when otherwise considered informative.

The mitochondrial DNA sequence information is from the entire mtDNA molecule (about 16.500 nucleotides). Alternatively, the information is obtained from a fragment hereof, not including the D-loop. The analysis of the complete mtDNA sequence or fragments thereof, may be carried out using existing technology for DNA sequencing. Analysis of the genetic markers may be based on DNA hybridisation assays (such as ASO hybridisation, DNA microchip, padlock), enzymatic cleavage assays (OLA, Taqman), enzymatic extension assays (minisequencing, pyrosequencing) or other suitable techniques for DNA typing.

In a second aspect, the present invention provides a kit for determining the geographic and population origin of a human, comprising means for sequence analysis of the complete human mitochondrial DNA of unknown origin and mitochondrial DNA sequence information of known origin. This sequence information can be provided as a leaflet with printed sequence information or as a printed reference to a database. The means for sequence analysis can be any means used in the known DNA sequencing procedures described above. The sequence of unknown origin is compared with the known sequences or a database

with complete mitochondrial genomes and conclusions about geographic and population origin can be made.

In an alternative embodiment, the kit comprises means for analysis of polymorphic sites, preferably located outside the D-loop. Polymorphic sites useful for the purposes of the present invention are listed in Table 1. The D-loop is numbered from 16028 to 577 (i.e. 16028 to 16569 and 1-577).

In an alternative emboiment, the present invention provides a specific set of laboratory reagents and conditions (protocol) for determining the type of nucleotide at a selected set of the polymorphic sites in the mitochondrial genome, based on using the pyrosequencing method.

Detailed description of the invention

The invention will now be described more closely in association with the accompanying drawings, in which

Figure 1 represents data matrices showing all informative nucleotide positions in 53 of the 124 individuals for which the complete mitochondrial genome has been determined, in decreasing order of frequency, in: a) the whole mtDNA genome, excluding the D-loop and; b) the D-loop. The trees on the left are cladograms with the same topology and numbering of individuals as the tree in Figure 2. Individuals of African decent are found exclusively below the dashed line and non-Africans above. The four major groups of sequences are boxed in blocks. The blocks denote groups of nucleotides identical in several sequences.

Figure 2 shows a Neighbor-Joining phylogram based on complete mtDNA genome sequences (but excluding the D-loop) of 53 of the 124 individuals examined, constructed using PAUP*4.0 Beta (Sinauer Associates) and bootstrapped with 1000 replicates (bootstrap values shown on nodes). The population origin of the individual is given at the twigs. The branches are shown block-wise as in Fig 1. Individuals of African descent are found exclusively below the dashed line and non-Africans above. The node marked '‡' refers to the MRCA of the youngest clade containing both African and non-African individuals.

Figure 3 shows mismatch distributions of pairwise nucleotide differences between 53 mtDNA genomes (excluding the D-loop) for: a) African and; b) non-African individual.

Figure 4 shows distribution in the number of differences in pairwise comparisons of 53 sequences for the D-loop only and for the complete mtDNA sequence.

Identification of genetic variation in the human mitochondrial genome

To assess the global genetic diversity in human mtDNA we have determined the complete mtDNA genome of a total of 124 individuals. Individuals were selected as to cover most of the genetic diversity in the human species.

The complete mtDNA genome was amplified in segments using PCR and the fragments sequenced. The primers used for PCR amplification were as described by Reider et al (1998). Sequencing was performed on the PCR products directly using BigDye (Applied Biosystems) chemistry. Separation of sequencing ladders was performed on the ABI 377 instrument for automated fragment analysis. Both forward and reverse strands were sequenced. Sequence analysis was performed using Sequencing Analysis 3.3 (Applied Biosystems) and sequence alignment was made with Sequencher 3.1.1 (Gene Codes).

All the 124 complete mtDNA sequences are unique. A total of 1122 polymorphisms were identified among these 124 individuals. A list of the sites, the types of alternative nucleotides found at each of these sites, and the frequency of these different nucleotides is shown in Table. 1.

Our study of the entire mitochondrial genome has significant distinctions from previous studies of the D-loop. Most importantly, the sequences outside of the D-loop evolve in an approximately 'clock-like' manner, enabling a more accurate measure of mutation rate, and therefore improved estimates of times to evolutionary events (Ingman et al. 2000). The difference between the D-loop and the remaining molecule is visually evident in the contrast between the jumbled arrangement of polymorphic sites in the D-loop and the clear haplotypes defined by the sites in the rest of the molecule (Fig. 1). The Neighbor–Joining tree constructed from our mtDNA sequences has a strongly supported basal branching pattern (Fig.

2). The three deepest branches lead exclusively to sub-Saharan mtDNAs, with the fourth branch containing both Africans and non-Africans. The deepest, statistically supported branch (NJ bootstrap = 100) provides compelling evidence of a human mtDNA origin in Africa. Our data shows that the genetic variation (polymorphism) in the mtDNA genome, outside the D-loop is subject to very little parallellism (i.e. very few independent substitutions in multiple mtDNA lineages at the same site), increasing the ability to determine the maternal lineage and deducing the geographic and population origin based on this information. while the D-loop has too many parallel substitutions at numerous sites to be useful for this purpose.

Utility of complete mtDNA genome variation for studies of human origin

Our analysis of the complete mitochondrial DNA genome (excluding the D-loop) have resulted in the following:

1) A very large number of novel polymorphisms outside the D-loop have been detected, many of which show a strong population-specific distribution. This increases the ability to determine the specific maternal lineage for an individual and define the geographic and population origin of an individual based on a biological sample.
2) A large database of complete mitochondrial genomes against which the genetic information from an individual can be compared, and geographic and population origin can be evaluated. For example, our database gives us an ability to identify with a very high probability whether the sample is obtained from an individual of African or Caucasian origin.

Kit for the analysis of a subset of mtDNA polymorphisms detected

The variation in mtDNA between individuals has been studied mainly using Sanger sequencing. Pyrosequencing is in comparison to many other techniques for genetic typing very quick, robust and easy to use (Ronaghi et al. 1996).

The kit developed for mtDNA determination in the present invention is based on the analysis of 10 PCR fragments covering highly informative sites in the entire mitochondrial genome (Table 2). The system is based on the analysis of 19 pyrosequencing reactions, including 4 HVI, 4 HVII and 11 coding region reactions. This allows the analysis of some of the most informative regions of the entire mtDNA. The method enables the analysis of the D-loop and the coding regions of the mtDNA and can used on a wide range of biological materials.

The analysis of the mitochondrial D-loop is performed from two separate PCR fragments. One hypervariable region I (HVI) fragment, which is analyzed in four separate pyrosequencing reactions. Similarly, the hypervariable region II (HVII) fragment is analyzed in four separate pyrosequencing reactions. In order to evaluate the technique, mitochondrial DNA (mtDNA) from a number of control samples have been analyzed. In total, 190 samples were analyzed for HVII, 120 samples were analyzed for HVI and finally 50 forensic forensic evidence materials were studied. The results were identical with sequencing data of the D-loop for the same individuals.

Pyrosequencing was further used to sequence 11 mtDNA coding regions in 36, previously sequenced, control samples. The 11 regions are chosen to cover the most informative sites throughout the entire mitochondrial genome using the previously generated complete mtDNA genome sequences. The results obtained using pyrosequencing were 100% identical to those obtained by the Sanger sequencing method.

To evaluate the system on limited amounts of DNA 50 biological samples with small amounts of DNA were analyzed for HVI and HVII by pyrosequencing. Among the materials analysed were samples from robber hoods, wigs, moustaches, shoes, cellular phones, watches, knives and guns. The pyrosequencing results were, for all the tested samples, identical to the results obtained using Sanger sequencing.

Combining the results of the eight HVI and HVII pyrosequencing reactions cover altogether approximately 88% (396/448) of the nucleotides after manual editing of the D-loop sequences. Complete Sanger sequencing of the D-loop gives a HVII fragment of 359 nucleotides and a HVI fragment of 403 nucleotides using the primer pairs L048/H408 and L15997/R16401 (Wilson et al. 1995). The pyrosequencing method covers 62% (222/359) of the nucleotides of the HVII fragment determined by Sanger sequencing and 56% (226/403) of the nucleotides of the HVI fragment determined by Sanger sequencing. In total, this gives a 59% coverage (448/762) of the D-loop determined by Sanger sequencing.

EXPERIMENTAL SECTION

Methodology for determining the complete mtDNA genome

The PCR

The mitochondrial genome is amplified in 24 overlapping fragments by the PCR technique, using 24 primer pairs, as described by Reider et al (1998). Both the forward and reverse strands of these 24 DNA templates are sequenced using established techniques for DNA sequencing such as theBigDye Primer Cycle Sequencing Ready Reaction (Applied Biosystems) kits diluted 1:1 with 1x sequencing buffer, using the following components

1uL of DNA template

4uL reaction mix (2uL kit and 2uL 1x sequencing buffer)

Extension reactions in MJ Research Tetrad thermocycler with programs as follows:

Forward reactions

(ramping at 1°/second)

96° 10 seconds

55° 5 seconds

70° 1 minute

<14 cycles>

96° 10 seconds

70° 1 minute

<14 cycles>

Hold at 4°

Reverse Reactions

(ramping at 1°/second)

96° 10 seconds

48° 5 seconds

70° 1 minute

<19 cycles>

96° 10 seconds

70° 1 minute

<14 cycles>

Hold at 4°

The extension products from these reactions are precipitated in 95% ethanol and the products spun down in a centrifuge at maximum speed for 25 minutes before removing the ethanol and allowing the pellets to air dry.

The pellets are rehydrated in a bromothymol loading buffer and loaded on an ABI 377 instrument for separation of sequencing ladders. A 96 well format was used allowing all the forward and reverse strands of 2 genomes to be run simultaneously. Sequencing gel analysis was performed using Sequencing Analysis 3.3 (Applied Biosystems) and the contigs were assembled into complete double stranded genomes with Sequencher 3.1.1 (Gene Codes). To aid in error checking, these completed genomes were then aligned with a consensus genome sequence of all the previously sequenced mtDNAs and all differences checked against the chromatograms for the new sequence. The new genome sequence could then be exported to the sequence database for analysis.

Sequence Analysis

The new complete mtDNA genome sequence is aligned with the other sequences in the reference database. Pairwise numbers of differences can be calculated against every sequence in the database, for instance using a computer program such as PAUP*. The resulting distance matrix is a list of the number of differences between the sequence under analysis and each reference sequence. From this information, the number of matches with 0, 1 & 2 differences between an individual sequence and those in the database can be can be calculated. In our database of 124 complete mtDNA sequences, the number of variable sites when considering the D-loop sequences only is much lower than that for the complete sequence.

Methodology and reagents for the kit assaying the mtDNA polymorphisms detected

DNA preparations

Human DNA was extracted from PBLs of 190 Swedish blood donors, to serve as control samples. The forensic evidence material was purified by one of three methods. The Wizard Genomic DNA Extraction Kit (Promega) was used to extract

DNA from most of the evidence material collected by cotton swabs and also from blood. Chelex 100 (Bio-Rad) was used to extract DNA from bloodstains. The method has been developed for extracting DNA from forensic-type samples for use with PCR (Walsh et al. 1991). Hair samples were extracted with an extraction procedure that uses proteinase K and DTT (Vigilant 1999).

Primer design and PCR

The software Primer Express® (Applied Biosystems) was used for PCR and sequence primer design. An optimal selection has been made to cover most of the informative polymorphism in the D-loop, where hypervariable region I and II (HVI and HVII) are PCR amplified in two different PCR products (Table 2). The coding region polymorphisms are amplified in 11 separate reactions and are analysed in 11 pyrosequencing reactions, using the forward PCR primer as sequencing primer. PCR amplification for control samples was set up in 70 μl containing 1.5 μl DNA, 200 nM of a normal and a biotinylated primer (Table 2), 200 nM of each dNTP, 1.5 mM $MgCl_2$, 2 U AmpliTaq® Gold DNA Polymerase and 1× GeneAmp® PCR Buffer II. The PCR amplification for the forensic evidence material was set up in 100 μl with 10 μl DNA, 200 nM of the normal and the biotinylated primer, 200 nM of each dNTP, 2.4 mM $MgCl_2$, 10 U AmpliTaq® Gold DNA Polymerase, 1.2× GeneAmp® PCR Buffer II, 0.16 mg/ml BSA and 10 % glycerol. The amplifications were performed in an ABI 9600 instrument (Applied Biosystems). The samples were kept for 10 min at 95°C followed by 45 cycles of 30 sec at 95°C, 45 sec at 60°C (53°C for the coding region primers) and 60 sec at 72°C. The final extension was lengthened to 7 min. Tubes that contained all PCR components, but without template (NTC), were used to ensure that the reagents were free of contamination.

Template preparation and pyrosequencing reaction

Streptavidin-coated beads were used as a solid phase support to obtain single-stranded biotinylated PCR products, as described by Pyrosequencing AB. Some reactions required 440 ng of Single Stranded DNA Binding Protein (SSB) (Amersham Pharmacia Biotech) added to the primed DNA template, prior to pyrosequencing (Ronaghi 2000). The sequencing was performed at room temperature and 15 μl of the 70 μl PCR reaction was used with 400 nM sequence primer. Enzyme and substrate mixture (prototype of PSQ™ 96 SQA Reagent Kit) were added to all samples in a PSQ™ 96 System (Pyrosequencing AB) with a

prototype of PSQ™ 96 SQA Software (sample entry, instrument control and evaluation). The procedure was carried out by stepwise elongation of the primer strand during sequential dispensation of different dNTPs (AαS, C, G and T) followed by degradation of nucleotides. Optimal cyclic dispensation orders were chosen for each fragment. The sequences were edited and compared with Anderson *et al.* reference standard (Anderson et al. 1981). The control samples were sequenced using an ABI 377 instrument and BigDye Terminator chemistry.

Table 1. List of nucleotide sites that are polymorphic in the 124 human mitochondrial genomes examined together with the alternative nucleotides found at each site and the frequency of this polymorphism in the 124 individuals. Sites indicated by an asterisc (*)have been described previously in conjunction with studied of mitochondrial disease. Numbering according to Andersson et al. (1981).

Position	Change	No.
43*	C ins	3
59	T-C	1
61	C-T	2
62	G-A	2
62	G-T	1
63*	T-C	2
64*	C-T	6
66*	G-A	2
70	G-A	1
72*	T-C	2
73*	A-G	108
89*	T-C	6
92	G-A	4
93*	A-G	7
95*	A-C	3
103*	G-A	2
114*	C-T	1
125*	T-C	3
127*	T-C	3
128	C-T	3
131	T-C	1
143*	G-A	7
146*	T-C	27
150*	C-T	17
151*	C-T	8
152*	T-C	29
153*	A-G	2
182*	C-T	6
183	A-G	2
185*	G-A	2
185*	G-T	1
186*	C-A	6
189*	A-G	6
189*	A-C	6
194*	C-T	4
195*	T-C	26
195	T-A	3
198*	C-T	5
199*	T-C	6
200*	A-G	4
204*	T-C	5
207*	G-A	4
208	T-C	2
211	A-G	1
212	T-C	5
215*	A-G	1
217*	T-C	1
225*	G-A	1
227*	A-G	1
227	A-T	1
228*	G-A	2
235*	A-G	1
236*	T-C	7
247*	G-A	12
248*	A del	7
249	A-G	1
250*	T-C	1
252	T-C	1
263*	A-G	121
271*	C-T	1
285	C-T	1
290*	AA del	2
291*	A del	1
295*	C-T	1
297*	A-G	6
303*	C ins	57
303*	CC ins	15
303*	C del	3
310*	T del	2
311*	C ins	121
316*	G-A	6
317	C-A	1
325*	C-T	1
339	A-G	1
357*	A-G	1
373*	A-G	1
374*	A-G	1
385*	A-G	2
408	T-A	1
453	T-C	2
456*	C-T	4
462*	C-T	1
464	A-G	1
465	C-T	1
467	C-T	2
471*	T-C	1
480	T-C	1
482	T-C	1
489*	T-C	30
499*	G-A	2
509	C-T	2
513*	G-A	3
514*	CACA ins	1
514*	CA ins	5
514*	CA del	37
542	C-T	1
547	A-G	1
548	C-T	1
571	C-T	1
574*	C ins	2
574*	CCC ins	3
574*	CCCCCC ins	3
574	A-G	1
591	C-A	1
593	T-C	3
629	T-C	1
663*	A-G	1
678	T-C	1
680	T-C	1
709*	G-A	11
710*	T-C	3
721	T-C	1
750*	G-A	1
753	A-C	1
769*	G-A	15
794	T-C	1
825*	T-A	12
827*	A-G	1
850	T-C	2
921	T-C	1
930*	G-A	1
942	A-G	1
1005	T-C	1
1007	G-A	1
1018*	G-A	15
1041*	A-G	1
1048	C-T	7
1119	T-C	3
1243	T-C	2
1375	C-T	2
1382*	A-C	1
1391	T-C	1
1397	T ins	1
1420*	T-C	1
1438*	A-G	120
1442	G-A	1
1503	G-A	1
1536*	A-G	1
1598*	G-A	3
1619	C-T	1
1628	C-T	1
1664*	G-A	2
1692	A-G	1
1719*	G-A	7
1736*	G-A	1
1738	T-C	1
1811*	A-G	2
1811	A-C	1
1834	T-C	1
1888*	G-A	4
1927	G-A	1
2000	C-T	1
2010	T-C	1
2056	G-A	1
2060	A-G	1
2092*	C-T	1
2156	A del	1
2157*	A ins	2
2218	C-T	1
2227	A del	4
2245*	A-G	2
2245	A-C	1
2251	A ins	3
2263	C-A	2
2283*	C-T	1
2308	A-G	2
2332	C-T	1
2352*	T-C	6
2359	C-T	1
2380	C-T	4
2394	A del	6
2416	T-C	5
2483	T-C	1
2503	A-C	1
2650	C-T	1
2706*	A-G	114
2755	A-G	2
2758*	G-A	12
2768	A-G	4
2789	C-T	2
2792	A-G	1
2836	C-A	2
2863	T-C	2
2885*	T-C	12
3010*	G-A	13
3027	T-C	3
3106*	C del	124
3159	T ins	1
3197*	T-C	3
3200	T-A	1
3203	A-G	2
3204	C-T	1
3206	C-T	1
3290*	T-C	1
3306	C-T	1
3308	T-C	1
3316	G-A	2
3351	C-A	1
3372	T-C	2
3381	A-G	1
3394	T-C	2
3396*	T-C	1
3398*	T-C	1

3399	A-T	1
3434*	A-G	1
3438*	G-A	4
3447*	A-G	1
3513	C-T	2
3516	C-A	4
3547	A-G	1
3549	C-T	1
3552	T-A	5
3591	G-A	1
3594	C-T	15
3645	T-C	3
3666	G-A	7
3693	G-A	1
3699	C-T	2
3705	G-A	1
3796	A-T	5
3798	C-T	1
3816*	A-G	2
3833	T-G	1
3843*	A-G	4
3866	T-C	1
3882	G-A	2
3918	G-A	1
3927	A-G	2
3970	C-T	4
3987	A-G	1
3990	C-T	1
3999	T-C	1
4015	C-T	1
4017	C-T	1
4023*	T-G	1
4047	T-C	1
4048	G-A	3
4050	C-T	1
4071	C-T	2
4086	C-T	1
4093	A-G	1
4104	A-G	15
4117*	T-C	9
4122	A-G	2
4164	A-G	2
4181	A-G	1
4191	A-G	1
4203	A-G	1
4216	T-C	5
4218	T-C	1
4230	C-T	1
4248*	T-C	1
4312*	C-T	5
4335	C-T	3

4336	T-C	1
4388	A-G	1
4452	T-C	1
4454	T-A	3
4454	T-C	2
4491*	G-A	2
4506	A-G	2
4508	C-T	1
4512	G-A	1
4529*	A-T	3
4541	G-A	2
4580*	G-A	1
4586*	T-C	5
4639*	T-C	1
4646*	T-C	1
4655*	G-A	1
4659*	G-A	1
4674	A-G	1
4688	T-C	3
4695	T-C	2
4700	C-A	1
4707	C-T	1
4715*	A-G	5
4742	T-C	1
4745	A-G	1
4769*	A-G	124
4793	A-G	1
4820	G-A	1
4824*	A-G	2
4833*	A-G	1
4834	C-T	1
4856	T-C	1
4883*	C-T	5
4892	C-T	1
4907	T-C	2
4914	C-G	1
4917	A-G	4
4973	T-C	1
4976	A-G	1
4977	T-C	1
4985*	G-A	124
4991*	G-A	2
4994	A-G	1
5004	T-C	1
5036	A-G	1
5046	G-A	1
5063	T-C	1
5075	T-C	1
5086	C-T	1
5090	T-C	1
5096	T-C	1

5105	T-C	1
5107	C-A	1
5108*	T-C	5
5147*	G-A	5
5153	A-G	1
5177	G-A	2
5178*	C-A	4
5206	C-T	1
5223	G-A	1
5230	T-A	1
5231*	G-A	3
5237	G-A	3
5255	C-T	1
5262	G-A	1
5263	C-T	4
5276	A-G	1
5285	A-G	1
5300	C-T	1
5302	T-C	2
5324	C-T	1
5330	C-A	1
5351*	A-G	2
5379	C-T	1
5393	T-C	1
5420	T-C	1
5426	T-C	1
5432	A-G	1
5442*	T-C	5
5460*	G-A	15
5465*	T-C	11
5471*	G-A	1
5483	T-C	1
5492	T-C	3
5562	T-C	1
5563	G-A	3
5580	T-C	1
5603*	C-T	3
5655	T-C	2
5656*	A-G	1
5711	A-G	2
5811*	A-G	2
5821*	G-A	1
5826	T-C	1
5836	A-G	1
5843	A-G	2
5894	A-C	1
5895*	C ins	1
5899	C del	2
5899	CCCCC ins	1
5951	A-G	6

5972	C-T	1
5984*	A-G	2
5987	C-T	1
5999*	T-C	1
6026	G-A	5
6047*	A-G	1
6071	T-C	6
6077	C-T	7
6083	C-T	1
6092	T-C	1
6104	C-T	1
6150	G-A	1
6164	C-T	2
6167	T-C	4
6182	G-A	2
6185	T-C	5
6221*	T-C	5
6242	C-T	1
6249	G-A	1
6253	T-C	2
6257	G-A	2
6260*	G-A	1
6296	C-A	1
6323	A-G	1
6324	G-A	1
6351	T-C	2
6366*	G-A	4
6371	C-T	1
6378	T-C	1
6392	T-C	3
6455*	C-T	2
6473	C-T	1
6524	T-C	1
6528	C-T	1
6548	C-T	1
6587*	C-T	1
6623	C-T	1
6632	T-C	1
6653	C-T	2
6680*	T-C	3
6719	T-C	12
6734*	G-A	4
6752	A-G	2
6755	G-A	3
6770	A-G	2
6827	T-C	1
6875	C-A	1
6881	A-G	1
6905	A-G	3
6917	G-A	1
6927	G-A	1

6929	A-G	1
6938	C-T	2
6962*	G-A	1
6989	A-G	1
7010	C-T	1
7028*	C-T	113
7055*	A-G	7
7058	T-A	1
7076	A-G	1
7146	A-G	11
7175	T-C	3
7196*	C-A	5
7202	A-G	2
7226	G-A	1
7256	C-T	15
7257	A-G	2
7268	T-C	1
7274	C-T	2
7298	A-C	1
7309	T-C	1
7319	T-C	1
7325	A-G	1
7337*	G-A	2
7340	G-A	3
7347	G-A	1
7388	A-G	1
7389	T-C	7
7403*	A-G	1
7419	G-A	1
7424*	A-G	1
7444	G-A	1
7493	C-T	1
7498	G-A	1
7521*	G-A	16
7561	T-C	1
7571	A-G	1
7581*	T-C	1
7598	G-A	2
7624	T-A	1
7645	T-C	1
7648	C-T	1
7657	T-C	1
7660	T-C	2
7673*	A-G	1
7675	C-T	1
7681	C-T	2
7684*	T-C	2
7693	C-T	2
7697	G-A	2
7711	T-C	1
7744	T-C	1

7768	A-G	2
7771	A-G	2
7805	G-A	1
7853*	G-A	2
7858*	C-T	1
7861	T-C	1
7867*	C-T	1
7948*	C-T	1
8014	A-G	1
8014	A-T	1
8020*	G-A	2
8027*	G-A	7
8075	G-A	1
8087	T-C	4
8149	A-G	1
8152*	G-A	2
8158	A-G	1
8206*	G-A	3
8248	A-G	1
8251*	G-A	6
8269*	G-A	2
8277	CCC ins	1
8277	CCCC ins	1
8277*	9bp del	17
8292*	G-A	2
8383	T-C	1
8387	G-A	1
8388	T-C	1
8389	A-G	1
8392	G-A	1
8404	T-C	8
8414*	C-T	4
8428*	C-T	3
8440	A-G	1
8460	A-G	2
8468*	C-T	12
8473*	T-C	1
8485	G-A	2
8491	A-G	1
8516	T-C	1
8525	A-G	1
8531	A-G	1
8533	G-A	1
8548	T-C	1
8562	C-T	1
8563*	A-G	1
8566	A-G	3
8572	G-A	2
8573	G-A	1
8575	C-T	2
8577	A-G	2

8584*	G-A	3
8603	T-C	2
8616	G-T	1
8618	T-C	1
8655	C-T	12
8697*	G-A	4
8701*	A-G	51
8703	C-T	1
8705	T-C	1
8711	A-G	1
8730	A-G	1
8733	T-C	1
8749	T-C	1
8764	G-A	2
8781	C-A	1
8784*	A-G	1
8790	G-A	9
8793*	T-C	2
8794*	C-T	1
8805	A-G	1
8818	C-T	1
8838	G-A	1
8841	C-T	1
8842	A-G	1
8842	A-C	1
8859	C-T	2
8860*	A-G	124
8865	G-A	1
8877	T-C	1
8910	C-T	1
8911	T-C	2
8928	T-C	2
8943	C-T	1
8958	C-T	1
8961	A-G	1
8964*	C-T	7
8973	A-G	1
8994*	G-A	2
9042	C-T	5
9053	G-A	2
9066	A-G	1
9072	A-G	6
9080	A-G	1
9103	T-C	4
9123*	G-A	11
9136	A-G	2
9140	C-T	3
9156	A-G	1
9158	C-T	1
9174	T-C	1
9221	A-G	3

9242	A-G	1
9248	C-T	1
9254	A-G	2
9257	C-T	1
9263	A-G	1
9266	G-A	2
9272	C-T	2
9288	A-G	1
9296*	C-T	1
9311	T-C	2
9347	A-G	5
9355	A-G	1
9356	C-T	1
9377*	A-G	1
9468	A-G	2
9477*	G-A	3
9509	T-C	2
9515	C-T	1
9522	C-T	1
9536*	C-T	1
9540*	T-C	51
9545	A-G	5
9548*	G-A	1
9554	G-A	1
9575*	G-A	1
9591	G-A	2
9599	C-T	1
9632	A-G	1
9647	T-C	2
9667*	A-G	2
9682	T-C	1
9755	G-A	4
9770	T-C	1
9776	C-T	1
9818	C-T	5
9824*	T-C	2
9824*	T-A	1
9852	A-G	1
9866	C-T	1
9899*	T-C	1
9932*	G-A	1
9938	T-C	2
9947*	G-A	1
9950*	T-C	1
9977	T-C	1
10034*	T-C	3
10088	C-T	1
10095	C-T	1
10097	A-C	2
10115	T-C	3
10118	T-C	6

10142	C-T	1
10186	C-T	1
10192	C-A	1
10197	G-A	1
10238*	T-C	13
10253	T-C	1
10310*	G-A	4
10313	A-G	1
10321	T-C	6
10358	A-G	1
10370*	T-C	1
10373	G-A	1
10398*	A-G	55
10400*	C-T	29
10410*	T-C	1
10427	G-A	1
10463*	T-C	5
10497	C-T	2
10499*	A-G	2
10505	T-C	1
10586*	G-A	6
10589*	G-A	5
10609	T-C	2
10640	T-C	1
10664	C-T	5
10667	T-C	1
10670	C-T	1
10685*	G-A	1
10688	G-A	13
10700*	A-G	2
10736	C-T	1
10786	T-C	2
10790	T-C	1
10792	A-G	1
10793	C-T	1
10801	G-A	1
10804	A-G	1
10810	T-C	13
10816	A-G	1
10819*	A-G	4
10873*	T-C	50
10876	A-G	2
10894	C-T	1
10914	G-A	2
10915*	T-C	8
10920	C-T	3
10933	C-G	1
10939	C-T	2
11016	G-A	3
11017*	T-C	2
11023	A-G	1

11150	G-A	1
11151	C-T	2
11167	A-G	2
11172	A-G	2
11176	G-A	4
11177	C-T	2
11204	T-C	1
11215	C-T	2
11251*	A-G	5
11257	C-T	2
11269	C-T	2
11288	C-T	2
11296	C-T	2
11299*	T-C	2
11332*	C-T	1
11335*	T-C	124
11339	T-C	1
11348	C-T	1
11365	T-C	1
11377*	G-A	1
11467*	A-G	5
11473	A-G	1
11476	C-T	1
11590	A-G	1
11611	G-A	1
11641	A-G	3
11647*	C-T	1
11653	A-G	3
11654*	A-G	1
11701	T-C	1
11719*	G-A	111
11722*	T-C	1
11807	A-G	1
11812*	A-G	2
11864	T-C	1
11884	A-G	1
11887	G-A	1
11893	A-G	1
11899	T-C	6
11909	A-G	1
11914*	G-A	16
11928	A-G	2
11944*	T-C	3
11959	A-G	1
11963*	G-A	3
11969	G-A	3
12007*	G-A	10
12026*	A-G	1
12049	C-T	1
12061	C-T	1
12070	G-A	3

12083	T-G	1
12121	T-C	3
12127	G-A	1
12134	T-C	1
12153	C-T	1
12172	A-G	2
12175	T-C	1
12192	G-A	1
12235	T-C	1
12236*	G-A	2
12239	C-T	10
12248	A-G	1
12308*	A-G	4
12338	T-C	1
12346	C-T	3
12358*	A-G	1
12361*	A-G	1
12366	A-G	1
12367	A-G	1
12372*	G-A	5
12379	C-T	1
12390	C-G	1
12397	A-G	1
12403	C-T	1
12405*	C-T	2
12406*	G-A	2
12414	T-C	1
12425	A-G	2
12477	T-C	1
12501*	G-A	2
12519	T-C	2
12530	A-G	1
12561	G-A	1
12603	C-T	1
12612*	A-G	1
12618	G-A	1
12630	G-A	3
12633*	C-A	2
12635	T-C	1
12672	A-G	1
12693	A-G	2
12696*	T-C	1
12705*	C-T	63
12720	A-G	3
12750	C-T	1
12768	A-G	2
12771*	G-A	1
12795*	G-A	1
12810*	A-G	6
12831	C-T	1
12842	T-C	1

12850	A-G	2
12879	T-C	2
12882	C-T	1
12930	A-T	2
12940	G-A	9
12957	T-C	1
12964	C-T	1
12973	C-T	1
12999	A-G	1
13020	T-C	4
13059	C-T	1
13065	C-T	1
13101	A-C	1
13104*	A-G	1
13105*	A-G	14
13135*	G-A	2
13143*	T-C	1
13145	G-A	1
13145	G-C	1
13149	A-G	1
13194	G-A	1
13197	C-T	1
13212	C-T	1
13221	A-G	1
13263*	A-G	5
13269	A-G	3
13276	A-G	5
13281*	T-C	3
13368*	G-A	5
13422	A-G	1
13452	C-T	1
13477*	G-A	1
13479	A-G	1
13485	A-G	6
13500	T-C	10
13506	C-T	12
13545	C-T	1
13590*	G-A	6
13594	A-G	1
13602	T-C	1
13611	A-G	1
13617*	T-C	3
13626	C-T	1
13641	T-C	4
13650	C-T	15
13651	A-G	3
13660	A-G	1
13681	A-G	2
13708	G-A	5
13722*	A-G	1
13759*	G-A	1

13780	A-G	2
13789	T-C	7
13803	A-G	2
13818	T-C	1
13819	T-C	2
13830	T-C	1
13846	C-A	1
13851	C-A	1
13880	C-A	1
13886	T-C	1
13927	A-T	1
13928	G-C	8
13934*	C-T	1
13958*	G-C	1
13965	T-C	1
13966*	A-G	1
13967*	C-T	1
13980	G-A	2
14000	T-A	6
14016*	G-A	1
14020	T-C	2
14022	A-G	10
14025	T-C	7
14034	T-C	3
14053	A-G	1
14062	A-T	1
14067	C-T	1
14070	A-G	1
14070	A-T	1
14088	T-C	7
14094	T-C	1
14097	C-T	2
14115	C-T	2
14118	A-G	1
14148	A-G	5
14152*	A-G	1
14178	T-C	7
14180*	T-C	1
14182*	T-C	5
14203	A-G	1
14209	A-G	4
14212*	T-C	4
14215	T-C	1
14230	A-G	1
14233*	A-G	1
14284	C-T	1
14308	T-C	5
14311	T-C	1
14318	T-C	5
14319	T-C	1
14338	C-T	3
14364	G-A	1
14371	T-C	2
14374	T-C	4
14384	G-C	1
14385	C-A	1
14404	C-T	1
14455	C-T	1
14470	T-C	2
14502*	T-C	4
14560*	G-A	7
14566	A-G	2
14569*	G-A	1
14571	T-A	1
14580	A-G	1
14587	A-G	1
14605	A-G	2
14620*	C-T	1
14668*	C-T	5
14687	A-G	1
14738	A-G	1
14766*	T-C	13
14767	T-C	1
14769*	A-G	1
14783*	T-C	29
14798*	T-C	1
14812	C-T	1
14862	C-T	1
14871	T-C	1
14890	A-G	2
14898	T del	1
14905*	G-A	6
14911*	C-T	6
14923	C-T	1
14946	C del	1
14947	C-T	1
14971	T-C	1
14979*	T-C	1
15016	C-T	1
15040*	C-T	1
15043*	G-A	31
15061	A-G	1
15077	G-A	1
15090	T-C	1
15110*	G-A	2
15113	A-G	1
15115*	T-C	2
15133	A-G	1
15136	C-T	3
15148	G-A	1
15172	G-A	3
15178	A-G	1
15191*	T-C	1
15204	T-C	5
15217*	G-A	1
15218*	A-G	2
15229	T-C	1
15244*	A-G	3
15258*	A-G	1
15300	T-C	2
15301*	G-A	39
15317*	G-A	2
15326*	A-G	124
15346	G-A	3
15358	A-G	1
15377	A-G	1
15378	T-C	1
15412	T-C	1
15431	G-A	5
15443	C-G	1
15452*	C-A	5
15454*	T-C	1
15470	T-C	1
15479	T-C	1
15487*	A-T	5
15498	G-A	1
15514	T-C	2
15516	C-T	1
15523	C-T	1
15535*	C-T	1
15607*	A-G	21
15613	A-G	1
15626	C-T	1
15629	T-C	1
15646	C-T	1
15663	T-C	4
15664	C-T	2
15670	T-C	4
15693*	T-C	1
15731	G-A	1
15746	A-G	10
15748	T-C	3
15758*	A-G	1
15784	T-C	4
15790	C-T	1
15803	G-A	1
15812	G-A	1
15833*	C-T	1
15849	C-T	1
15852	T-C	2
15884*	G-C	1
15885	C-T	2
15889	T-C	1
15904*	C-T	1
15912	C-T	1
15924*	A-G	7
15927*	G-A	1
15928*	G-A	4
15930	G-A	1
15932	T-C	1
15937	A-T	3
15941	T-C	1
15942	T-C	1
15944	T del	1
15954*	A-C	1
15970	T-C	1
16038	A-G	1
16051*	A-G	4
16067*	C-T	1
16069*	C-T	1
16071*	C-T	1
16075*	T-C	1
16086*	T-C	3
16093*	T-C	6
16108*	C-T	1
16111*	C-T	4
16124*	T-C	1
16126*	T-C	7
16129*	G-A	24
16140*	T-C	3
16144*	T-C	6
16145*	G-A	6
16147*	C-T	3
16148*	C-T	11
16153*	G-A	2
16162*	A-G	2
16163*	A-G	1
16166*	A-C	2
16168*	C-T	1
16169*	C-T	1
16169*	C-A	1
16170	A-C	1
16171	A-G	1
16172*	T-C	12
16175	A-G	1
16176*	C-T	5
16182*	AA del	13
16183*	A del	11
16184	A ins	2
16184*	C ins	9
16184*	CC ins	6
16184*	CCC ins	7
16184	CA ins	2
16184*	C del	4

16184*	C-T	2
16186*	C-T	1
16187*	C-T	14
16188*	C-T	2
16188	C-A	2
16188*	C-G	1
16189*	T-C	44
16192	C-T	7
16193*	C-T	2
16207*	A-G	1
16209*	T-C	7
16213*	G-A	5
16214*	C-T	4
16217*	T-C	14
16218*	C-T	3
16221*	C-T	1
16221	C-A	1
16320*	C-T	6

16222*	C-T	1
16223*	C-T	58
16224*	T-C	1
16230*	A-G	5
16234*	C-T	5
16235*	A-G	7
16241*	A-G	9
16242*	C-T	2
16242	C-A	1
16243*	C-T	1
16245*	C-T	1
16247*	A-G	8
16249	T-C	5
16255	G-A	2
16256	C-T	4
16257*	C-T	2
16258	A-C	1

16261*	C-T	11
16263*	T-C	1
16265*	A-C	7
16265*	A-T	1
16265*	A-G	1
16266*	C-T	6
16268	C-T	1
16270*	C-T	7
16271	T-C	3
16274*	G-A	8
16278	C-T	21
16284	A-G	1
16286*	C-T	1
16286	C-G	1
16287*	C-T	4
16288	T-C	1
16290*	C-T	1

16291	C-G	2
16291	C-T	6
16292	C-T	2
16293*	A-G	4
16294*	C-T	15
16296*	C-T	2
16297*	T-C	3
16298	T-C	7
16303	G-A	2
16304	T-C	7
16309*	A-G	2
16311	T-C	37
16312	A-G	1
16316	A-G	1
16317	A-G	2
16318	A-T	1
16319	G-A	4

16324*	T-C	3
16325*	T-C	4
16326	A-G	1
16327	C-T	6
16335*	A-G	1
16337	C-T	1
16343*	A-G	6
16344*	C-T	1
16352*	T-C	1
16354*	C-T	2
16355*	C-T	1
16356*	T-C	4
16357*	T-C	9
16360	C-T	8
16362*	T-C	19
16366	C-T	1
16368*	T-C	1
16371	A-G	1
16390*	G-A	8
16391	G-A	2
16399	A-G	5
16400*	C-T	1
16428	G-A	1
16438	G-A	3
16439	C-A	1
16468	T-C	1
16483*	G-A	1
16488	C-T	1
16497	A-G	2
16509	T-C	1
16519*	T-C	69
16526	G-A	2
16527	C-T	3

Table 2. PCR and sequencing primers used for the analysis of mtDNA polymorphism.

Primer Name*	PCR Primer	Sequene Primer	T_m (°C)	Sequence (5´-3´)	Dispensa-tion order	SSB Required
II 45	X	X	51,3	ATG CAT TTG GTA TTT TCG TCT G	TCGA	
II 111		X	43,3	ACC CTA TGT CGC AGT ATC T	TCGA	
II 162		X	50,1	CGC ACC TAC GTT CAA TAT TAC A	CTGA	
II 216		X	44,3	TTA ATG CTT GTA GGA CAT AAT AA	CTGA	
II 287 B	X		51,2	TTG TTA TGA TGT CTG TGT GGA AAG		
C 431	X	X	53,6	CAC CCC CCA ACT AAC ACA	ACGT	
C 637 B	X		53,1	GTG ATG TGA GCC CGT CTA A		
C 2988	X	X	51,8	CGA TGT TGG ATC AGG ACA	ACGT	
C 3216 B	X		52,4	GG TGG GTG TGG GTA TAA		
C 3403	X	X	56,8	CTA CGC AAA GGC CCC AA	CAGT	X
C 3641 B	X		56	GCT AGG CTA GAG GTG GCT AGA A		
C 4156	X	X	52,5	CAA CTC ATA CAC CTC CTA TGA AA	ACGT	
C 4367 B	X		52,6	TTG GAT TCT CAG GGA TGG		
C 4882	X	X	49	CCA TCT CAA TCA TAT ACC AAA	ATGC	
C 5138 B	X		50	GGA GTT TAA GTT GAG TAG TAG GAA		
C 8665	X	X	52	CAA TGA CTA ATC AAA CTA ACC TCA	ATGC	
C 8803 B	X		51,1	TAA ATG AGT GAG GCA GGA GT		
C 12346	X	X	47,9	CAC ACT ACT ATA ACC ACC CTA A	ACGT	X
C 12541 B	X		49,1	CTC AGT GTC AGT TCG AGA TAA		
C 12673	X	X	45,7	AAC ATT AAT CAG TTC TTC AAA	ACGT	X
C 12861 B	X		46,7	GTT GTA TAG GAT TGC TTG AA		
C 14747	X	X	53,4	ATG ACC CCA ATA CGC AAA	ACGT	X
C 14949 B	X		54,1	TGG GCG ATT GAT GAA AAG		
C 15883	X	X	50,2	GGC CTG TCC TTG TAG TAT AAA	ACGT	
C 16083 B	X		52	GGT TGT TGA TGG GTG AGT C		
C 16496	X	X	48,5	GAC ATC TGG TTC CTA CTT CA	ACGT	X
C149 B	X		47,7	ATG AGG CAG GAA TCA AA		
I 16105	X	X	50,2	TGC CAG CCA CCA TGA ATA	CTGA	X
I 16168		X	45,1	CCA ATC CAC ATC AAA ACC	CTGA	X
I 16203		X	40,4	AGC AAG TAC AGC AAT CAA	CTGA	
I 16266		X	42,1	CCC ACT AGG ATA CCA ACA	CTGA	
I 16348 B	X		51,8	GAC TGT AAT GTG CTA TGT ACG GTA AA		

*I = HVI, II = HVII, C = coding region, B = biotin labeled reverse primer.

References

Allen, M. Engström, AS. Meyers, S. Handt, O. Saldeen, T. von Haeseler, A. Paabo, S. Gyllensten, U. 1998. Mitochondrial DNA sequencing of shed hairs and saliva on robbery caps: sensitivity and matching probabilities. Journal of Forensic Science 43(3):453-64.

Anderson, S., A. T. Bankier, et al. (1981). Sequence and organization of the human mitochondrial genome. Nature 290(5806): 457-65.

Bodenhagen D, Clayton DA. The number of mitochondrial deoxyribonucleic acid genomes in mouse L and human HeLa cells. J Biol Chem 1974;249:7991-5.

Brown WM, Prager EM, Wang A, Wilson AC. Mitochondrial DNA sequences of primates: tempo and mode of evolution. J. Mol Evol 1982;18: 225-39.

Gill P, Ivanov PL, Kimpton C, Piercy R, Benson N, Tully G, et al. Identification of the remains of the Romanov family by DNA analysis. Nat Genet 1994;6:130-5.

Ginther C, Issel-Tarver L, King MC. Identifying individuals by sequencing mitochondrial DNA from teeth. Nat Genet 1992;2:135-8.

Hagelberg E, Sykes B. Ancient bone DNA amplified. Nature 1989;342:485.

Handt O, Richards M, Trommsdorff M, Kilger C, Simanainen J, Georgiev O, et al. Molecular genetic analyses of the Tyrolean ice man. Science 1994;264:17758.

Higuchi R, von Berolding CH, Sensabaugh GF., Erlich HA. DNA typing from single hairs, Nature, 1988;332:543-6.

Holland MM, Fisher DL, Mitchell LG, Rodriquez WC, Canik JJ, Merril CR, et al. Mitochondrial DNA sequence analysis of human skeletal remains: Identification of remains from the Vietnam war. J For Sci 1993;38:542-53.

Hutchison III CA, Newbold JE, Potter SS, Edgell M. Maternal inheritance of mammalian mitochondrial DNA. Nature 1974;251:536-538.

Ingman, M., H. Kaessmann, et al. (2000). "Mitochondrial genome variation and the origin of modern humans." Nature 408(6813): 708-13.

Pääbo S, Higuchi RG, Wilson AC. Ancient DNA and the polymerase chain reaction J Biol Chem 1989;264:9709-12.

Reider, MJ Taylor, SL, Tobe, VO & Nickerson, DA. Automating the identification of DNA variations using quality-based fluorescence re-sequencing: analysis of the human mitochondrial genome. Nucleic Acids Res. 26, 967-973 (1998).

Ronaghi, M., S. Karamohamed, et al. (1996). "Real-time DNA sequencing using detection of pyrophosphate release." Anal Biochem 242(1): 84-9.

Stoneking M, Hedgecook D, Higuchi RG, Vigilant L, Erlich HA. Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. Am J Hum Genet 1991;48: 370-82.

Vigilant, L. (1999). "An evaluation of techniques for the extraction and amplification of DNA from naturally shed hairs." Biol Chem 380(11): 1329-31.

Walsh, P. S., D. A. Metzger, et al. (1991). "Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material." Biotechniques 10(4): 506-13.

Wilson, M. R., D. Polanskey, et al. (1995). "Extraction, PCR amplification and sequencing of mitochondrial DNA from human hair shafts." Biotechniques 18(4): 662-9.

Wilson MR, Holland MM, Stoneking M, DiZinno, JA, Budowle B. Guidelines for the use of mitochondrial DNA sequencing in forensic medicine. Crime Lab Dig 1993;20:68-77.

CLAIMS

1. A method for determining the geographic or population origin of a human comprising the following steps:
a) determining the complete nucleic acid sequence of the mitochondrial genome including polymorphic sites in a sample from a human subject who's origin or identity is studied; and
b) relating the information from step a) to mitochondrial nucleic acid sequence information of known origin.

2. A method according to claim 1, wherein the known information in step b) is derived from a database of nucleic acid sequence information from humans of diverse origin.

3. A method according to claims 1 or 2, wherein the mitochondrial nucleic acid sequence is the complete nucleic acid sequence of the mtDNA genome, excluding the D-loop.

4. A method according to claims 1 or 2, wherein the mitochondrial nucleic acid sequence comprises the polymorphic sites mentioned in Table 1.

5. A method according to any of the above claims, wherein the human mitochondrial nucleic acid sequences is determined by DNA sequencing, or in the case of genetic markers, on assays such as DNA hybridisation assays (ASO, SSO hybridisation, DNA microchip, padlock), enzymatic ligation assays (OLA, padlock) enzymatic cleavage assays (Taqman), enzymatic extension assays (minisequencing, pyrosequencing) or other assays for typing of genetic polymorphisms.

6. A method according to claim 5, wherein the mitochondrial nucleic acid sequence is determined by pyrosequencing.

7. A kit for determining geographic or population origin of a human, comprising means for analysis of a selected set of the genetic markers (polymorphism) from the sites listed in Table 1 of the present patent application.

8. A kit according to claim 7, wherein the selected set of genetic markers are outside the D-loop.

9. A kit according to claim 7 or 8, wherein the means for analysis are selected from the reagents listed in Table 2 of the present patent application.

AMENDED CLAIMS

[received by the International Bureau on 22 January 2002 (22.01.02);
original claims 1-9 replaced by amended claims 1-13 (2 pages)]

AMENDED CLAIMS under Article 19, PCT

1. A method for determining the origin or identity of a human comprising the following steps:

a) determining polymorphic sites in the complete nucleic acid sequence of the mitochondrial genome in a sample from a human subject who's origin or identity is studied; and

b) relating the information from step a) to mitochondrial nucleic acid sequence information of known origin.

2. A method according to claim 1, wherein the known information in step b) is derived from a database of nucleic acid sequence information from humans of diverse origin.

3. A method according to claims 1 or 2, wherein the mitochondrial nucleic acid sequence is the complete nucleic acid sequence of the mtDNA genome, excluding the D-loop.

4. A method according to claims 1 or 2, wherein the mitochondrial nucleic acid sequence comprises the polymorphic sites mentioned in Table 1.

5. A method according to any of the above claims, wherein the human mitochondrial nucleic acid sequences is determined by DNA sequencing, or in the case of genetic markers, on assays such as DNA hybridisation assays (ASO, SSO hybridisation, DNA microchip, padlock), enzymatic ligation assays (OLA, padlock) enzymatic cleavage assays (Taqman), enzymatic extension assays (minisequencing, pyrosequencing) or other assays for typing of genetic polymorphisms.

6. A method according to claim 5, wherein the mitochondrial nucleic acid sequence is determined by pyrosequencing.

7. A method according to any of the above claims, wherein the means for analysis are selected from the reagents listed in Table 2 of the present patent application.

8. A kit for determining origin or identity of a human, comprising means for analysis covering informative sites in the entire mitochondrial genome.

9. A kit according to claim 8, wherein the informative sites are outside the D-loop.

10. A kit according to claims 8 or 9, wherein the means for analysis are selected from the reagents listed in Table 2 of the present patent application.

11. A kit according to any of the claims 8-10, wherein the means for analysis are amplifying primers, sequencing primers and means for detection of polymorphism.

12. A kit according to any of the claims 8-11, wherein the means for analysis are means for DNA sequencing, or in the case of genetic markers, on assays such as DNA hybridisation assays (ASO, SSO hybridisation, DNA microchip, padlock), enzymatic ligation assays (OLA, padlock) enzymatic cleavage assays (Taqman), enzymatic extension assays (minisequencing, pyrosequencing) or other assays for typing of genetic polymorphisms.

13. A kit according to any of the claims 8-12, wherein the means for analysis are pyrosequencing means.

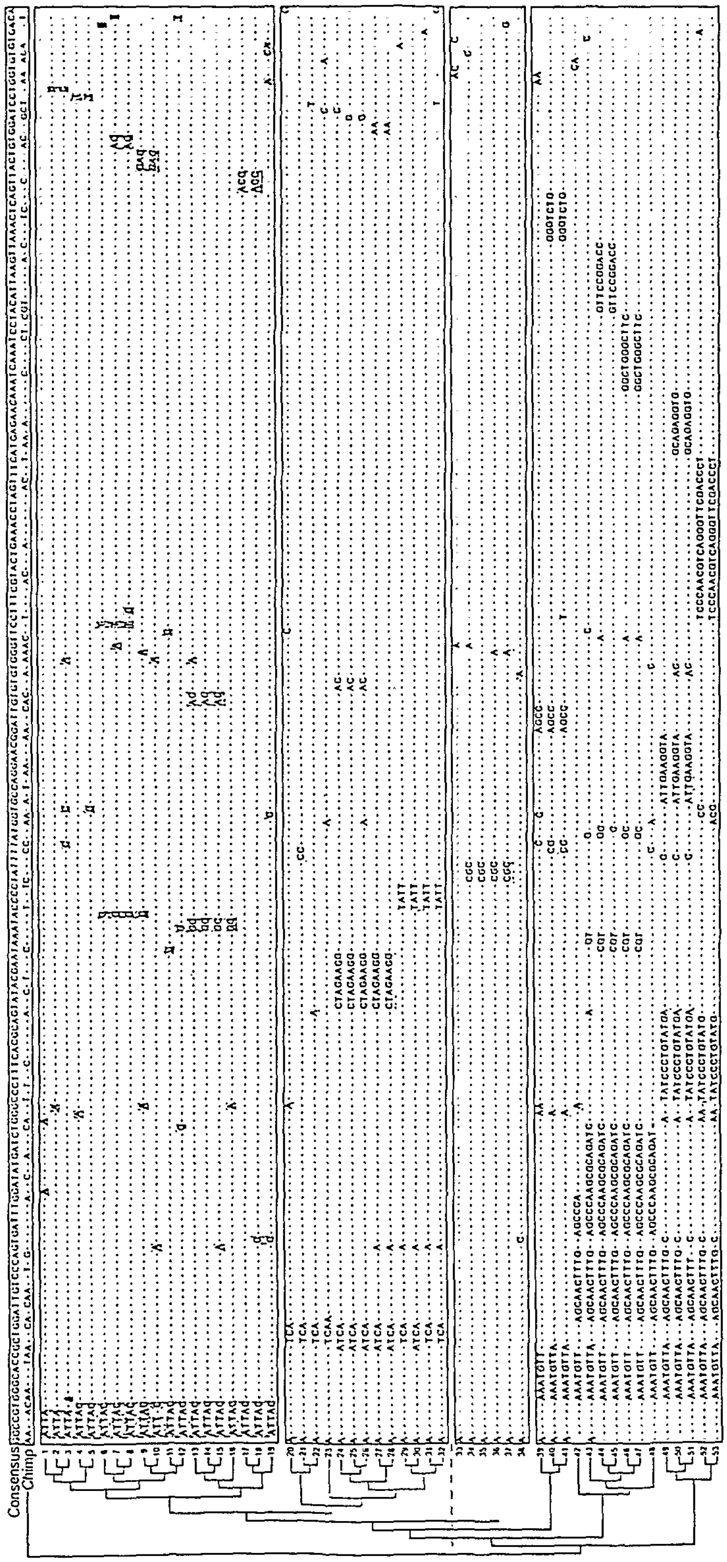


FIG. 1A

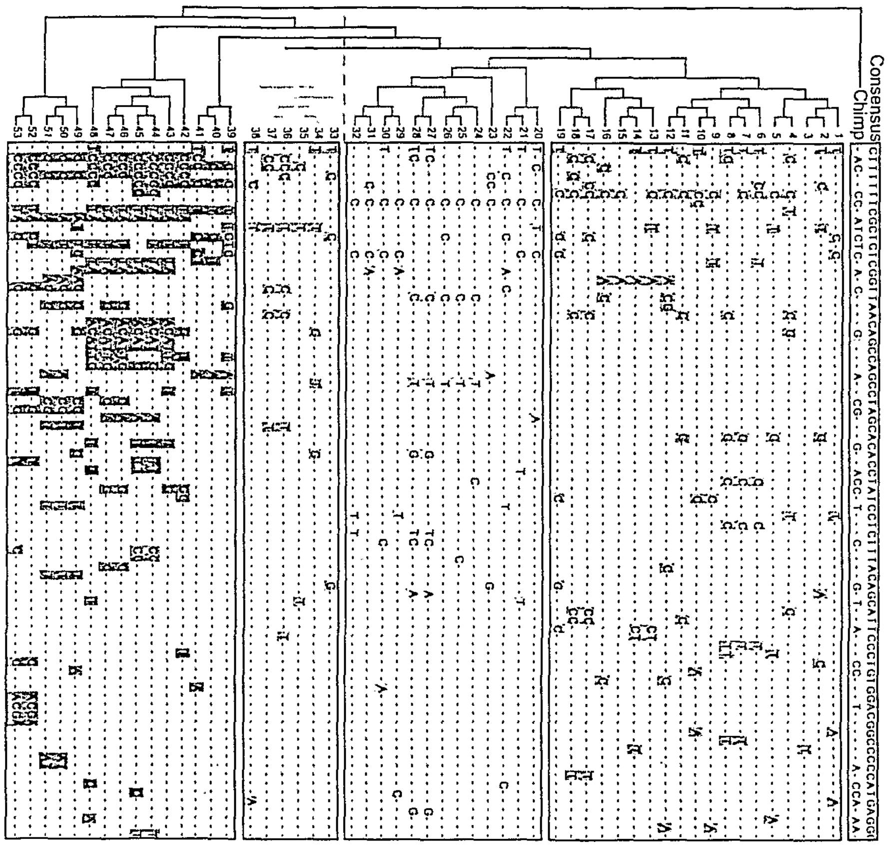


FIG. 1B

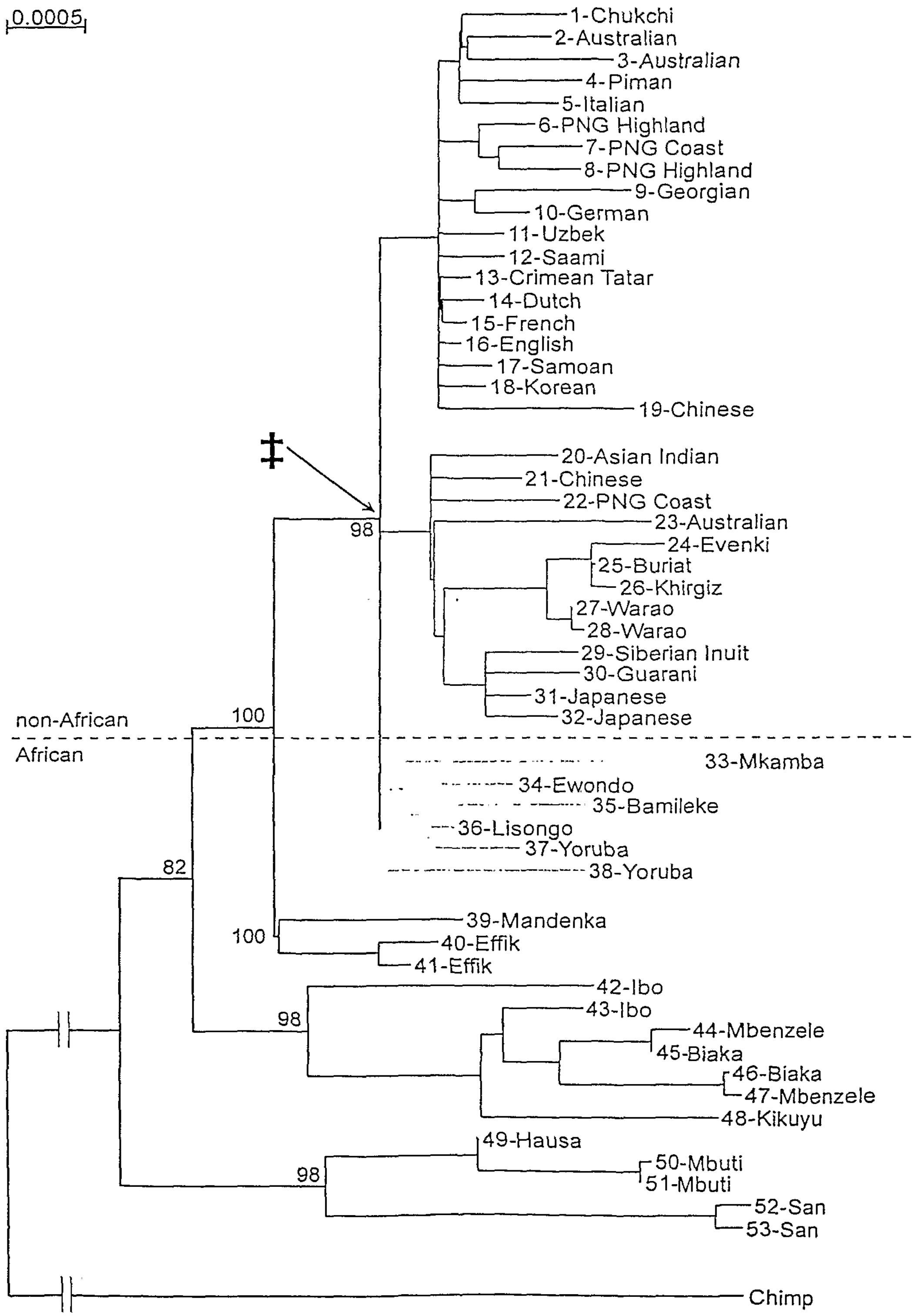

0.0005
1-Chukchi
2-Australian
3-Australian
4-Piman
5-Italian
6-PNG Highland
7-PNG Coast
8-PNG Highland
9-Georgian
10-German
11-Uzbek
12-Saami
13-Crimean Tatar
14-Dutch
15-French
16-English
17-Samoan
18-Korean
19-Chinese
‡
20-Asian Indian
21-Chinese
22-PNG Coast
23-Australian
24-Evenki
25-Buriat
26-Khirgiz
27-Warao
28-Warao
29-Siberian Inuit
30-Guarani
31-Japanese
32-Japanese
98
100
non-African
African
33-Mkamba
34-Ewondo
35-Bamileke
36-Lisongo
37-Yoruba
38-Yoruba
82
39-Mandenka
100
40-Effik
41-Effik
42-Ibo
98
43-Ibo
44-Mbenzele
45-Biaka
46-Biaka
47-Mbenzele
48-Kikuyu
49-Hausa
98
50-Mbuti
51-Mbuti
52-San
53-San
Chimp

FIG. 2

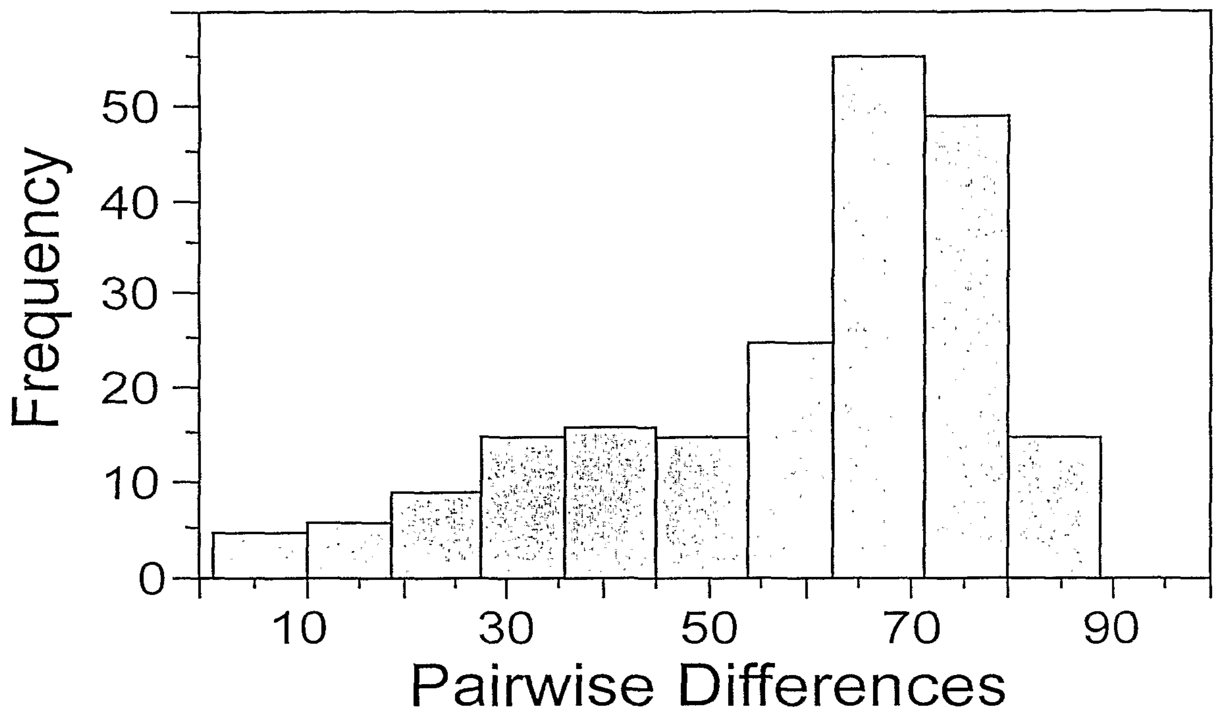


FIG. 3A

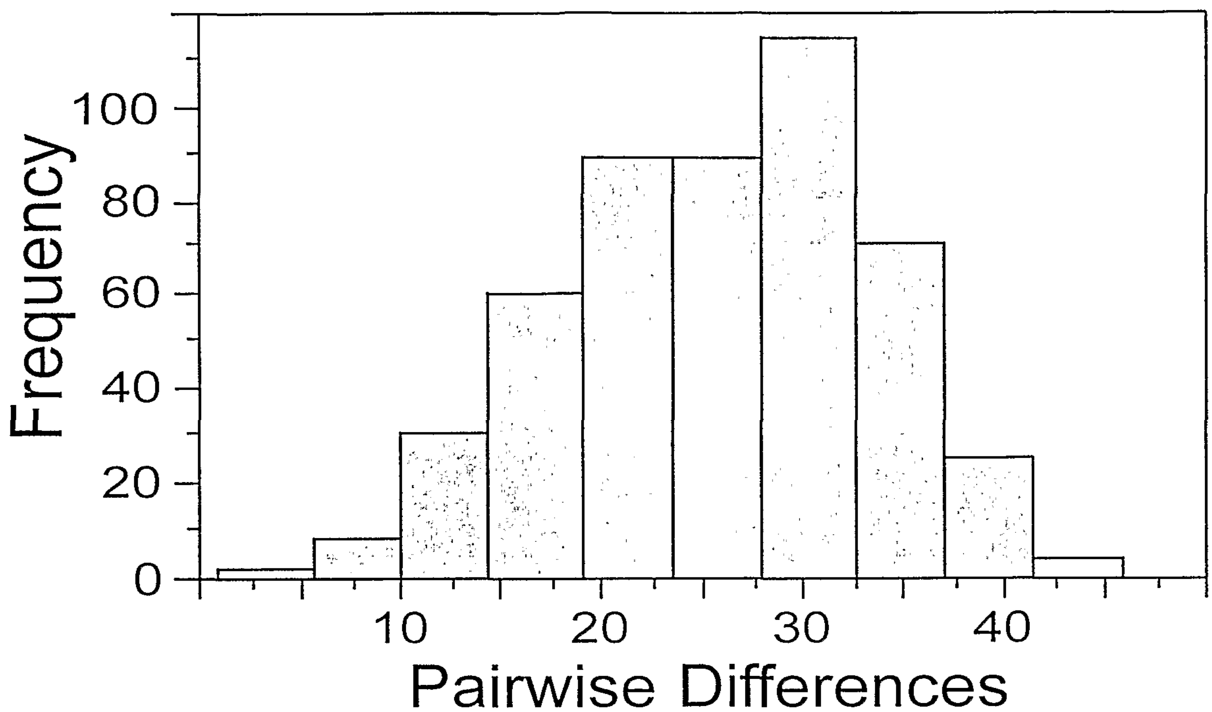
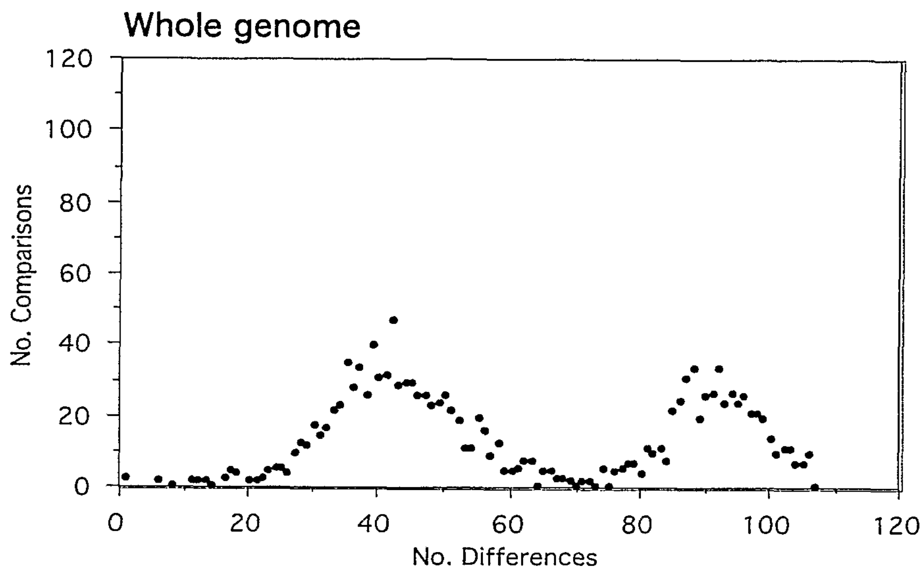


FIG. 3B


Whole genome
120
100
80
60
40
20
0
No. Comparisons
0
20
40
60
80
100
120
No. Differences

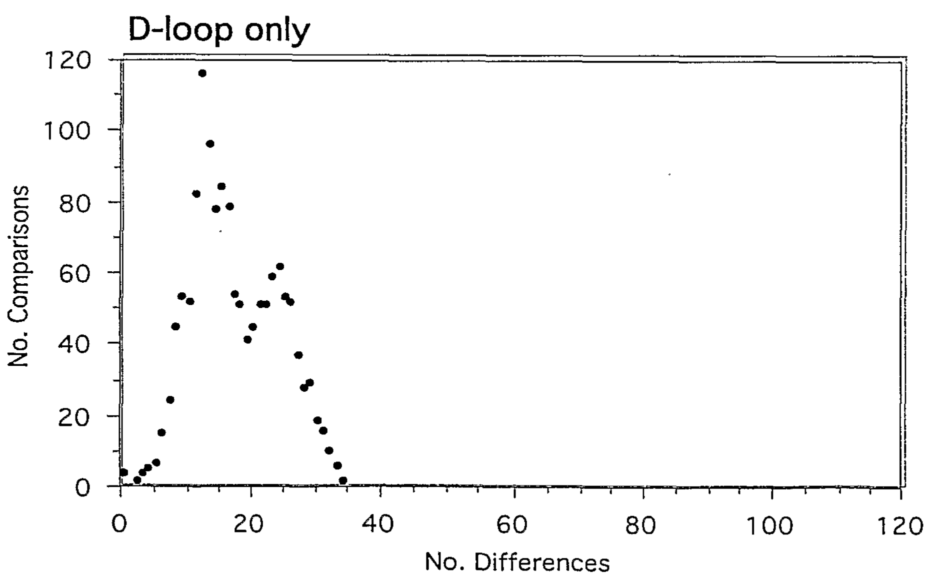

D-loop only
120
100
80
60
40
20
0
No. Comparisons
0
20
40
60
80
100
120
No. Differences

FIG. 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SE 01/01691

A. CLASSIFICATION OF SUBJECT MATTER

IPC7: C12Q 1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC7: C12Q, C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

SE,DK,FI,NO classes as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-INTERNAL, WPI DATA, BIOSIS, MEDLINE, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Am.J.Hum.Genet, Volume 66, 2000, Saara Finnilä et al, "Phylogenetic Network of the mtDNA Haplogroup U in Northern Finland Based on Sequence Analysis of the Complete Coding Region by Conformation-Sensitive Gel Electrophoresis", page 1017 - page 1026, see page 1025, column 1, line 11 - line 20 --	1-9
Y	Proc.Natl.Acad.Sci., Volume 88, March 1991, Anna Di Rienzo et al, "Branching pattern in the evolutionary tree for human mitochondrial DNA", page 1597 - page 1601, see page 1597, column 1, line 14 - line 42 --	1-9

[X] Further documents are listed in the continuation of Box C. [X] See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
20 November 2001	22 -11- 2001
Name and mailing address of the ISA/ Swedish Patent Office Box 5055, S-102 42 STOCKHOLM Facsimile No. +46 8 666 02 86	Authorized officer Lars Wallentin/ELY Telephone No. +46 8 782 25 00

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	EP 0812922 A2 (AFFYMETRIX, INC.), 17 December 1997 (17.12.97), page 2, line 24 - line 43; page 5, line 26 - line 35	1-9

--

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SE01/01691

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: **7-9**
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

See extra sheet

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest ☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

The subject matter of claims 7-9 fails to comply with the clarity and conciseness requirements of PCT Article 6 to such an extent that a meaningful search on the basis of these claims is impossible for the following reason:

- The expression "means for analysis of a selected set of the genetic markers…" is unclear and does not define a searchable scope, e.g. the number of possible combinations that evolves when applying the different markers listed in Table 1 is almost unlimited and lack any link to a corresponding human population.

Further, the description does not contain examples on how to use the kit or how to conclude about geographic or population origin.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SE 01/01691

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0812922 A2	17/12/97	EP 0902885 A	24/03/99
		JP 10099085 A	21/04/98
		JP 2000512744 T	26/09/00
		US 5981956 A	09/11/99
		US 6207960 B	27/03/01
		US 2001030290 A	18/10/01
		WO 9743611 A	20/11/97